DOCUMENT RESUME

ED 214 876 SP 019 813

AUTHOR Sirotnik, Kenneth A.; And Others
TITLE Some Methodological Issues in Developing Measures of Classroom Learning Environment. A Study of Schooling in the United States. Technical Report Series, No. 6.
INSTITUTION California Univ., Los Angeles. Graduate School of Education
SPONS AGENCY Institute for Development of Educational Activities, Dayton, Ohio.
PUB DATE 80
NOTE 21p.

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Attitude Measures; *Classroom Environment; Factor Analysis; *Item Analysis; *Research Design; Secondary Education; Secondary School Students; Student Attitudes; *Test Construction; Test Interpretation; Test Items; *Test Validity

ABSTRACT

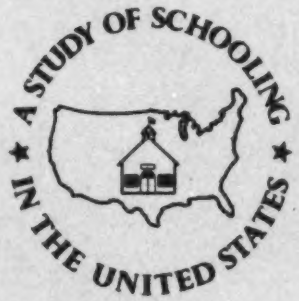
Methodological and conceptual issues arising from a study of student perceptions of secondary school class climate are discussed. Through a pilot study, 113 Likert-type items were collected, consisting of teacher-focused dimensions, classroom-focused dimensions, and instructional practices. Factor and cluster analyses were used at the item level. Three correlation matrices were examined: students as individual units, students as units within classes, and classes as units. Two levels of schooling, junior and senior high school, were also computed. Efforts were focused on finding similarities between correlation matrices, and 18 class climate series appeared to be useful for further study: student competitiveness; student decision making; peer esteem; student cliqueness; classroom dissonance; task difficulty; student apathy; organization; student compliance; classroom physical appearance; student satisfaction; knowledge of results; teacher clarity; teacher favoritism; teacher authoritarianism; teacher punitiveness; teacher enthusiasm; and teacher concern. The choice of unit items for analysis was critical to the outcome of the study, and differences will be further explored at the scale level. (FG)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED214876

A STUDY OF SCHOOLING

Technical Report Series



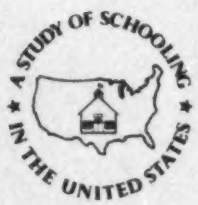
"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
John I. Goodlad
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)"

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
✓ This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

Distributed by the Laboratory in
School and Community Education

Graduate School of Education
University of California
Los Angeles, California 90024

SP 019 813



SUPPORTED BY

The Danforth Foundation
International Paper Company Foundation
The JDR 3rd Fund
Martha Holden Jennings Foundation
Charles F. Kettering Foundation
Lilly Endowment, Inc.
Charles Stewart Mott Foundation
National Institute of Education
The Needmor Fund
Pedamorphosis, Inc.
The Rockefeller Foundation
The Spencer Foundation
United States Office of Education

Conducted under the auspices of the Institute for
Development of Educational Activities (/I/D/E/A/)
A Division of the Charles F. Kettering Foundation

John I. Goodlad, Principal Investigator

SOME METHODOLOGICAL ISSUES IN DEVELOPING MEASURES OF CLASSROOM LEARNING ENVIRONMENT

Kenneth A. Sirotnik
Mitchell A. Nides
Gerald A. Engstrom

Technical Report No. 6

1980

A Study of Schooling is based upon the assumption that improving schools requires knowing what is happening in and around them. A comprehensive data-base of contextual information was obtained from students, teachers, administrators, parents and observers at all grade levels in thirty-eight elementary and secondary purposively sampled schools. *It is strongly recommended that readers of any technical report in this series first read Technical Report No. 1 which outlines the details, scope and limitations of the Study as a whole.*

It must be understood that this series of technical reports does not constitute the Study. Some reports are highly specific "molecular" inquiries while others take a more "molar" view across data sources, schooling levels, etc. Some reports are more methodological in nature arising out of issues in data analysis. Many of the reports quite naturally overlap in data analysed and interpretations rendered. Some authors have approached their task as consisting mostly of data description with little discussion beyond the presentation of the data. Others have ventured further into the realm of interpretation and speculation. *It must be further understood that data-based inferences can and do differ among researchers who come at the data from differing points-of-view.* Authors, therefore, are duly acknowledged for each report and are responsible for the material presented therein.

Some Methodological Issues in Developing Measures
of Classroom Learning Environment

Introduction

As one of the many substantive components of A Study of Schooling, student perceptions of their classroom climate or learning environment constituted a major research focus.[1] Many questions have been formulated over the course of the Study concerning the relationships (if any) among dimensions of classroom learning environment as perceived by students, teachers, and trained classroom observers.

These and other questions are currently serving as guides to what is essentially an exploratory data analysis of all the information collected from the 38 schools in our Study during the Spring and Fall of 1977. The 13 elementary schools and 25 secondary schools (12 junior high or middle schools and 13 senior high schools) in this sample have yielded student questionnaire, teacher interview and questionnaire, and observation data for over 1000 classes (526, 369, and 134 at the senior high, junior high, and elementary levels, respectively). All students present in classes over the several days of questionnairing were included in the sample resulting in over 20,000 individual surveys (11,051; 9,106; and 3,444 at the senior high, junior high, and elementary levels, respectively). Classes were sampled at each school as randomly as possible, such that two classes per grade were obtained at each elementary school and samples of all subject areas and tracking levels were included in the classes at secondary schools.

A major portion of the class-specific section of the student surveys consisted of attitudinal, Likert-type items designed to measure class climate

(27-item, 43-item, and 113-item versions at the early elementary, upper elementary, and secondary levels, respectively). We wish to focus on the secondary version of the instrument in this report, and specifically, on the conceptual and methodological issues arising during the development and analysis of the 113 items.

Instrument Development

Early on in the design phase of the Study, we decided (for reasons of efficiency and comparability of data and data collection) to construct only three student questionnaire booklets in optical scan format (early and upper elementary and secondary forms). Thus, the secondary class climate survey had to be functional over the entire 7-12 grade span, primarily in terms of reading level. Additionally, substantive interests on the /I/D/E/A/ staff ranged from the highly phenomenological views students might have about their classroom experience to the more environmental views students might have about classroom processes. No single existing instrument appeared to meet this spectrum of interests. Nor was there any survey which contained items written at a level of sufficiently low reading ability to accommodate our grade span. The staff, therefore, set out to develop a class climate instrument.

Actually, two such instruments emerged, representing the work of the two conceptual viewpoints noted above. One instrument focused heavily on the student's interpersonal needs as they might interact with those of the particular teacher and remaining students in a given classroom. (See Novotney, 1975 and Novotney and Overman, 1975.) The items and emergent constructs for this instrument had a kind of "psychological" bent to them written mostly in first person singular--item scales with titles like "teacher concern," "peer

esteem," and "teacher punitiveness" were the result. The second instrument focused more heavily on classroom process, having items more "sociological" in nature and written mostly in third person plural. The resulting instrument, largely adapted from Gordon's (1963) work, contained item scales with titles like "task," "authority," "compliance," and "order."

After a major pilot study and extensive factor analysis, it was determined that some of the constructs in both instruments overlapped, and that a single, comprehensive survey encompassing both perspectives would be most useful. Although still no single instrument existed which we felt met our needs, several good surveys existed from which we borrowed or modified constructs and items. (We relied primarily on the LEI instrument developed by Anderson, 1971.) Adding these to the items we already had, and augmenting the collection with new items pertaining to "instructional practices" constructs which we felt were not sufficiently emphasized in our items and those of others, a pool of 113 items (written at approximately 7-8th grade reading level) and 29 constructs resulted. The left side of Table 1 lists these

---

Insert Table 1 about here

---

constructs and an exemplary item for each. (It should be noted that the first four constructs are the result of factor analytic pilot studies; the remaining constructs are essentially a priori.) In an effort to place what we have developed into perspective given currently available climate inventories, our survey might loosely be described as having combined some of the teacher-focused dimensions found in the CES (Moos and Trickett, 1974) and the classroom-focused dimensions found in the LEI (Anderson, 1971) with our brand of more personalized items and items focused more directly on instructional practices such as appropriate practice and knowledge of results.

Scale Analysis: Methods

At this point we departed considerably from what has been done by others in the field in two important ways: (1) we explored the viability of our a priori constructs by using factor (and cluster) analysis at the item level and (2) we computed more than one kind of correlation matrix for this purpose. These modes of analysis suggested themselves as a result of our acknowledgement of a critical, and conceptually, very sticky issue having to do with appropriate units of analysis. In short, more than one correlation coefficient can be computed between two variables when the data can be organized into subsets determined by one (or more) grouping factors.[2]

Thus, for example, the correlation between any two items on the climate inventory can be computed using students as units irrespective of their class membership (the "total" analysis); students as units, but within each class separately and then averaged across classes (the "pooled within" analysis); and/or the classes themselves as units using item means based on the students within the classes (the "between" analysis). Adding level of schooling (junior vs. senior) to these possibilities, the following six correlation matrices could be (and have been) computed for our 113-item survey.[3]

	Units of Analysis		
Correlation Matrix Type	All Secondary...	High School...	Junior High...
Total	Students (N=20,157)	Students (N=11,051)	Students (N=9,106)
Between	Classes (N=895)	Classes (N=526)	Classes (N=369)
Pooled Within	Students (N=20,157)	Students (N=11,051)	Students (N=9,106)

Why bother with all these different matrices? The answer is that the values in each matrix can be quite different and lead to substantially dissimilar item structures, i.e., scales as derived from cluster (or factor) analyses. Even more importantly, these potential differences are not statistical—they are conceptual. Choosing between one or another type of correlation coefficient amounts to selecting one or another perspective on the role of an item as a measurement device.

The psychometric implications of selecting different units to be analyzed are quite complex, and they are only beginning to be explored by relatively few researchers. The reader is referred to Sirotnik(1979) for a more in-depth discussion and illustration of the problem. Essentially, if the property being measured is conceptualized as fundamentally systemic, then the between analysis is most appropriate. The averaging process across individuals within groups becomes nothing more than an operational device for arriving at a single systemic measurement. If the property is viewed as fundamentally phenomenological, given a group or systemic context, then the within analysis is most appropriate. In this case, although the group mean can still be a relevant statistic, it must be conceptualized as an average of individual perceptions with its meaning being derived from the meanings of those individual perceptions. The total analysis becomes relevant only when grouping factors become irrelevant.

Obviously, in assessing class climate, grouping factors (namely, classrooms) are relevant. We thus set out to interpret the dimensionality of both pooled within and between correlation matrices. Before turning to the results of these analyses, it is important to note our methodological approach, one which has usually been labeled rational scale development.

In effect, we do not rely exclusively on a priori scalings, nor do we strictly adhere to any one empirical structure generated by factor (or cluster) analysis. Instead, we combine both approaches hoping to arrive at a set of scales with some empirical and substantive utility. This approach necessarily results in correlated scales, which corresponds quite nicely with our view on the operation of climate constructs.

Scale Analysis: Results

Since we are presently in the throes of psychometric analysis of the class climate inventory, only an outline of some preliminary findings will be presented here.

Item Variance Components

Correlation ratios were computed for each item to determine the relative percentages of between and within class variation.[4] In no case did between class differences account for more than 35% of the item's total variation; and, in fact, most items never surpassed 25%.

These results might be rather disappointing to one who believes that consensus among students within a classroom is necessary to characterize the class as exhibiting a climate or an "environmental press." Certainly these results have a lot to do with item substance and form, i.e., whether they are high or low inference in structure and whether they are written in personalized or depersonalized forms. Generally, items with larger correlation ratios tend to tap structural or teacher characteristics and/or to be written in other than the first person singular. Those with lower ratios tend to tap into student characteristics and inter-student relationships and to be written in the first person singular.

Nevertheless, our items are not unlike those of other climate surveys, and in view of these findings, they would have to be labeled moderate to high inference items. And this is probably the inevitable result of writing items at levels of abstraction beyond, for example, an item like: "The color of this room is mostly purple." Yet, even relatively high inference items can yield between class variance in sufficient amounts to offer the potential for interpretable between class item structures.

Item Dimensionality

In light of our discussion and the above results, analyzing the between correlation matrix represents an attempt to identify scales based on average student perceptions which could be useful in making predictions at the class level. Analyzing the pooled within correlation matrix, then, represents an attempt to identify scales based on individual student perceptions which could have predictive utility at the individual student level within the context of their own classes.

Since both measurement prespectives are relevant to our study and each can yield quite different results, we are currently exploring the structures of all the correlation matrices noted above. Given the many different substantive (and analytic) interests of the staff, we necessarily focused our efforts on finding similarities between the various correlation matrices. And, to date, we have arrived at 18 dimensions (using 90 of the initial 113 items) which are similar for the between and pooled within analyses and appear to be useful for further analyses in the Study. These 18 class climate or learning environment scales are listed on the right side of Table 1 along with their alpha coefficients of internal consistency, computed at

both the pooled within and between levels of analysis. Note that of the original 29 constructs, 14 (scales 1-14) remained pretty much intact while four "hybrids" emerged as new dimensions. In effect, several scales such as Rules and Regulations, Appropriate Practice, and Materials dropped out while scales such as Teacher Condescension, Teacher Flexibility, and Student Freedom merged together forming a more complex construct.

Of course, the necessary expediency of finding common dimensions for the Study had the effect of minimizing the differences that obtain between the different units of analysis and the different schooling levels. Nonetheless, differences occur, are not trivial, and have the potential for interesting substantive interpretation--a process we are only just beginning. Although the pooled within and between analyses show small clusters approximating the 18 dimensions of our instrument, they also show larger clusters which differ in their grouping of these 18 dimensions.

Table 2 illustrates the differences in higher order clustering that

---

Insert Table 2 about here

---

obtain between the different schooling levels and units of analysis. The numbers inside the table are merely labels identifying eleven clusters which run more or less in common across the various analyses. For example, the pooled within analysis at the high school level shows a large composite dimension (cluster #1) consisting of teacher concern, favoritism and clarity, student satisfaction and competitiveness, classroom physical environment, and the instructional practices scales "knowledge of results" and, partially, "organization." (This corresponds to the same general "class affect" factor that emerges in a six-factor rotated solution of the pooled within high school correlation matrix.)

In this paper, we will refrain from trying to follow up on all the intriguing points of dissimilarity between the various analyses. The most obvious generalizations that can be made are, first, that there are substantial differences between the within and between analyses and, next, that the two levels (high and junior high) are identical for the within analysis but differ greatly for the between. Apparently, secondary students as individuals may perceive the dynamics of classroom environment similarly, regardless of grade level. But at the classroom level, constructs emerging based upon average item perceptions can be quite different at the two schooling levels.

Briefly, it would appear from the between analyses that cluster 1 is basically the same for the high school and junior high school classes but the similarity ends with this cluster. The "highs" have another relatively large bipolar cluster containing most of the instructional practices items and the student satisfaction, student apathy and teacher clarity items. The remaining scales each form separate clusters. The "junior" analysis, by contrast, shows evidence of two relatively large clusters beyond the first. These two clusters, containing scales related to classroom instruction and student satisfaction respectively, appear to separate into clusters according to whether the items in the scales are positively or negatively worded.

Concluding Remarks

Clearly, the choice of which unit (and schooling level) to analyze during instrument development is a critical one. While it would be tempting to ignore these differences between the various analyses, to do so creates the risk of misinterpreting associations between scales. To be sure, we have found support for many of the existing constructs of classroom learning environment; but we have also found that they can correlate in very different

ways depending upon the measurement perspective taken at the item level. We expect to further explore the substance of these differences using second order cluster (and factor) analyses at the scale level.

Moreover, the unit-of-analysis focus has also made it more difficult to ignore other methodological problems which have also been largely ignored in attitude scale construction. A good example is that of positive versus negative item wording, which we briefly alluded to in the last paragraph of the preceding section. As do most researchers, we included both positively and negatively worded items pertaining to the same (we thought) construct in the original item pool as a control for response bias, particularly acquiescence. We have since discovered very strong tendencies for clusters to form around the direction of item wording, perhaps to the exclusion of item content. Unfortunately, we cannot isolate the dynamics of this tendency since we did not neatly balance the variables of item direction, grammatical form, substance and level of inference across all items in the instrument. It would appear, however, that there may be a lot to be learned from the psycholinguistic aspects of attitudinal items and the way in which these facets may interact with respondent and contextual characteristics.[5]

Not surprisingly, the operation of this "response bias/substance/item form" variable is different depending on the unit of analysis. For example, an individual response style interpretation might argue for the positive-negative item clustering to be evident in the pooled within analysis, but be "averaged out" in the between analysis. Indeed, such is almost the case, since cluster number 1 from the within analysis consists of the scales containing the positively worded teacher and instructional practices items, the student satisfaction and compliance items and the positive physical environment items. Cluster number 2, in contrast, consists of the scales

containing the negative teacher and instructional practices items and the student apathy items. Furthermore, in the between analysis, the high school classes tend to combine both positive and negative scales into the same cluster. Yet, the between analysis for junior high classes shows perhaps the clearest separation of positive and negative clustering. Whatever is operating to produce these tendencies, it is clearly multifaceted.

Our hope is that as further analysis and interpretation proceed, further clarification on these and other methodological concerns will be achieved. Our expectations, however, are more realistic; data-based inquiries of this sort always seem to end up raising a lot more questions than answers.

Notes

[1]More information on the background and current activities of A Study of Schooling can be found in the series of four articles to be published in the Phi Delta Kappan beginning with Goodlad, Sirotnik and Overman (1979).

[2]A number of sources are available for the reader interested in following up on the "unit-of-analysis" issue. The reviews and discussion contained in Cronbach (1976) and Knapp (1977) are good starting points. See also the section on "analytic models" in Walberg (1976).

[3]Including schools (within levels) as a factor in the design would probably be advisable; but we felt the situation was already sufficiently complex both analytically and interpretively and we, therefore, ignored this additional grouping factor.

[4]The correlation ratio (or eta squared) is the proportion of between variation in the total variation using an ordinary analysis of variance breakdown of the item's variance between and within classes. Computational formulas for this and the various correlation and alpha coefficients referred to here can be found in Knapp (1977) and Sirotnik (1979).

[5]A limited review of the class climate (and self-concept) literature has uncovered a number of factor analytic studies producing positive and negative factors, each containing items which appear to be merely opposite ends of a single, bipolar dimension. We intend to expand this review and revisit the literature on response sets (and styles) in an effort to further clarify this issue.

References

Anderson, G.J. The assessment of learning environments: a manual for the learning environment inventory and the my class inventory. Halifax, Nova Scotia: Atlantic Institute of Education, 1971.

Cronbach, L.J. (with assistance of Deken, J.E. and Webb, N.). Research on classrooms and schools: Formulation of questions, design and analysis. Occasional Paper of the Stanford Evaluation Consortium, Stanford University, 1976.

Goodlad, J.I., Sirotnik, K.A., and Overman, B.C. An overview of 'A Study of Schooling.' Phi Delta Kappan, 1979, 61, 174-178.

Gordon, C.W., and Adler, L.M. Dimensions of teacher leadership in classroom social systems. Los Angeles: University of California, 1963.

Knapp, T.R. The unit-of-analysis problem in application of simple correlation analysis to educational research. Journal of Educational Statistics, 1977, 2, 171-186.

Moos, R.H., and Trickett, E.J. Manual for the classroom environment scale. Palo Alto, CA.: Consulting Psychologists Press, 1974.

Novotney, J.M. A framework for analysis of classroom affect. Paper presented at the American Educational Research Association, 1975.

Novotney, J.M., and Overman, B.C. Measuring classroom affect. Paper presented at the American Educational Research Association, 1975.

Sirotnik, K.A. Psychometric implications of the unit-of-analysis "problem" (with examples from the measurement of organizational climate). Technical report (July, 1979). A Study of Schooling, Los Angeles: Institute for Development of Educational Activities, Inc.

Walberg, H.J. Psychology of learning environments: Behavioral, structured, or perceptual? In Shulman, L.S. (Ed.), Review of Research in Education, 1976. Illinois: F.E. Peacock, 1977.

Table 1

A Priori and Rationally Derived
Learning Environment Constructs
With Item Examples and Alpha
Coefficients of Internal Consistency

A Priori Construct Titles	Sample Item[a]	Rationally Derived Construct Titles	No. of Items	Alpha Coefficients[b]	
				Pooled Within	Between
Teacher Concern	The teacher listens to me.	Teacher Concern	8	.85	.96
Teacher Punitiveness	The teacher punishes me unfairly.	Teacher Punitiveness	6	.74	.87
Peer Esteem	I like working with other students in this class.	Peer Esteem	7	.77	.89
Classroom Dissonance	Students in this class yell at each other.	Classroom Dissonance	3	.64	.93
Teacher Enthusiasm	This teacher seems to like being a teacher.	Teacher Enthusiasm	3	.62	.86
Teacher Favoritism	This teacher likes some students in this class better than others.	Teacher Favoritism	3	.61	.81
Teacher Clarity	The teacher gives clear directions.	Teacher Clarity	4	.65	.88
Student Competitiveness	There is a lot of competition in this class.	Student Competitiveness	4	.65	.83
Student Cliqueness	Certain students stick together in small groups.	Student Cliqueness	3	.47	.76
Student Satisfaction	Students feel good about what happens in this class.	Student Satisfaction	4	.69	.90
Student Compliance	The students in this class usually do the work assigned.	Student Compliance	4	.67	.85
Student Apathy	Students don't care about what goes on in this class.	Student Apathy	4	.49	.86

a. 4-point item response format used: strongly and mildly agree and disagree.

b. Pooled within and between alpha coefficients are computed using pooled within and between covariance matrices respectively. It can be shown mathematically that whenever class differences exists, between item correlations (and thus alpha coefficients) will always be greater than those for the pooled within analysis. In view of the conceptual differences in approach, this does not mean that between scales are necessarily "better than" pooled within scales.

Table 1 (Cont'd.)

A Priori Construct Titles	Sample Item	Rationally Derived Construct Titles	No. of Items	Alpha Coefficients: Pooled Within	Alpha Coefficients: Between
Classroom Physical Appearance	I like the way this classroom looks.	Classroom Physical Appearance	2	.62	.86
Knowledge of Results	The teacher tells us how to correct the mistakes in our work.	Knowledge of Results	4	.71	.90
Task Difficulty	Some of the things the teacher wants us to learn are just too hard.	Task Difficulty	4	.62	.83
Time (Pacing/Speed)	The teacher gives me too much work to do in this class.				
Teacher Condescension	This teacher treats us like children.	Teacher Authoritarianism	8	.78	.95
Teacher Flexibility	This teacher will never admit when he/she is wrong.				
Student Freedom	We don't feel like we have any freedom in this class.				
Student Decision-Making	Students help decide what we do in this class.	Student Decision-Making	8	.64	.88
Individualization	Different students can do different things in this class.				
Organization	Things are well planned in this class.	Organization	11	.76	.91
Perceived Purpose	Our teacher gives us good reasons for learning in this class.				
Goals and Objectives	Students know the goals of this class.				
Grading	The grades or marks I get in this class have nothing to do with what I really know.				
Teacher Task Behavior	We know exactly what we have to get done in this class.				

Table 1 (Cont'd.)

A Priori Construct Titles	Sample Item	Rationally Derived Construct Titles	No. of Items	Alpha Coefficents: Pooled Within	Alpha Coefficents: Between
Rules and Regulations	Everyone in this class knows what we may or may not do.	These items did not produce, or fit into, any rationally derived constructs.			
Appropriate Practice	We get to practice what we learn in this class.				
Materials	I can always find books or materials in this class that are interesting to me.				

Table 2

Higher Order Clustering Tendencies
Among the Items of the 18 Learning Environment Scales[a]

		Teacher Concern	Teacher Enthusiasm	Teacher Punitiveness	Teacher Authoritarianism	Teacher Favoritism	Teacher Clarity	Knowledge of Results	Student Satisfaction	Classroom Physical Appearance	Student Compliance	Organization	Student Apathy	Task Difficulty	Classroom Dissonance	Student Cliqueness	Student Decision-Making	Peer Esteem	Student Competitiveness
Pooled Within	Senior High School Students	1	1	2	2	2	1	1	1	1	1	1/2	2	2	3	3	4	5	6
	Junior High School Students	1	1	2	2	2	1	1	1	1	1	1/2	2	2	3	3	4	5	6
Between	Senior High School Classes	1	1	1	1	7	2	2	2	11	8	2	2	9	3	10	4	5	6
	Junior High School Classes	1	1	1/3	1	3	2	1	2	2	2	3	3	3	3	6	4	5	5

a. Reading across indicates which scales form the larger clusters, as numbered inside the table. Scales with two numbers (e.g., 1/3) have items which split into two different clusters.